

# 確率・統計基礎: 点推定、十分統計量

森 立平

# 設定

- ▶  $x \in \mathcal{X}$ : データ、知っているもの。
- ▶  $\theta \in \Theta$ : パラメータ、母数、知りたいもの、 $\Theta \subseteq \mathbb{R}$  とか  $\Theta \subseteq \mathbb{R}^n$  とか
- ▶  $p(x | \theta)$  or  $\Pr(X = x | \theta)$ : 尤度 (条件付き確率密度関数 or 条件付き確率質量関数)
- ▶  $p(\theta)$ : 事前確率密度関数、仮定する場合もある (ベイズ推定) が今回は仮定しない。

## 推定したいものの例

$$g: \Theta \rightarrow \mathbb{R}$$

▶  $g(\theta) = \theta.$

▶  $g(\theta) = \mathbb{E}[X \mid \theta].$

▶  $g(\theta) = \mathbb{V}[X \mid \theta].$

▶  $g(\theta) = \mathbb{E}[s(X) \mid \theta] \quad \text{for some } s: \mathcal{X} \rightarrow \mathbb{R}.$

# 損失関数とリスク関数

## Definition (損失関数、リスク関数)

以下を満たす関数  $L: \Theta \times \mathbb{R} \rightarrow \mathbb{R}$  を  $g(\theta)$  の損失関数という。

▶ 非負性

$$L(\theta, d) \geq 0 \quad \forall \theta \in \Theta, d \in \mathbb{R}.$$

▶ 正しいときはゼロ

$$L(\theta, g(\theta)) = 0 \quad \forall \theta \in \Theta.$$

また、対応するリスク関数  $R: \Theta \times (\mathcal{X} \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$  を以下で定義する。

$$R(\theta, \delta) := \mathbb{E}[L(\theta, \delta(X)) \mid \theta].$$

この期待値は  $X$  に関して取るものであるが、 $\delta$  がランダムである場合はそのランダムネスについても期待値を取る。

# 損失関数の例

- ▶ **0-1 損失関数**  $L(\theta, d) = \mathbb{1}_{\{d \neq g(\theta)\}}$ . このとき、  
 $R(\theta, \delta) = \Pr(\delta(X) \neq g(\theta) \mid \theta)$
- ▶ **二乗誤差損失**  $L(\theta, d) = (g(\theta) - d)^2$ . このとき、  
 $R(\theta, \delta) = \mathbb{E}[(g(\theta) - \delta(X))^2 \mid \theta]$ . この  $R(\theta, \delta)$  を **平均二乗誤差** (MSE; mean squared error) という。

## 連続確率変数の条件付確率

$Y$  が離散確率変数の場合、

$$\Pr(X \in A \mid Y = y) = \frac{\Pr(X \in A, Y = y)}{\Pr(Y = y)}$$

$Y$  が連続確率変数の場合、

$$\Pr(X \in A \mid Y = y) := \lim_{\epsilon \rightarrow 0} \frac{\Pr(X \in A, Y \in [y, y + \epsilon])}{\Pr(Y \in [y, y + \epsilon])}$$

# 十分統計量

## Definition (十分統計量)

$\mathcal{X}$  上の確率分布族  $p_\theta$  について  $T: \mathcal{X} \rightarrow \mathcal{Y}$  が十分統計量  $\stackrel{\text{def}}{\iff}$  任意の  $A \subseteq \mathcal{X}$  と  $t \in \mathcal{Y}$  について

$$\Pr(X \in A \mid T(X) = t, \theta)$$

が  $\theta$  に依存しない。

「 $T(X)$  は  $X$  に含まれている  $\theta$  の情報をすべて含む」というのが直感的な意味。

基本的には  $X$  の余計な部分を削ぎ落して  $\theta$  の推定に必要な部分だけを取り出したものを  $T(X)$  としたい。

$T(x) = x$  は自明な十分統計量であるが、余計な部分を削ぎ落していないので意味がない。

# 十分統計量は推定において十分

## Theorem

任意の推定関数  $\delta: \mathcal{X} \rightarrow \mathbb{R}$  と十分統計量  $T: \mathcal{X} \rightarrow \mathcal{Y}$  について、あるランダムな  $\eta: \mathcal{Y} \rightarrow \mathbb{R}$  が存在して、推定関数  $\eta \circ T$  は同じリスク関数の値をとる。

## Proof.

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}[L(\theta, \delta(X)) \mid \theta] \\ &= \sum_x \Pr(X = x \mid \theta) L(\theta, \delta(x)) \\ &= \sum_x \Pr(X = x, T(X) = T(x) \mid \theta) L(\theta, \delta(x)) \\ &= \sum_x \Pr(X = x \mid T(X) = T(x), \theta) \Pr(T(X) = T(x) \mid \theta) L(\theta, \delta(x)) \\ &= \sum_x \Pr(X = x \mid T(X) = T(x)) \Pr(T(X) = T(x) \mid \theta) L(\theta, \delta(x)) \\ &= \sum_t \Pr(T(X) = t \mid \theta) \sum_{x \in T^{-1}(t)} \Pr(X = x \mid T(X) = t) L(\theta, \delta(x)). \end{aligned}$$

$$\eta(t) = \delta(x) \quad \text{with probability } \Pr(X = x \mid T(X) = t)$$

とおけば同じリスク関数の値になる。



## 十分統計量の例 1/3

$X_1, \dots, X_N$  が独立同分布なベルヌーイ分布  $\text{Ber}(p)$

$$\Pr(X_1 = x_1, \dots, X_N = x_N \mid p) = p^{x_1 + \dots + x_N} (1 - p)^{N - (x_1 + \dots + x_N)}$$

に従っているとき、 $T(x_1, \dots, x_N) = \sum_{k=1}^N x_k$  は  $p$  の十分統計量である。  
実際、

$$\begin{aligned}\Pr(T(X_1, \dots, X_N) = t \mid p) &= \Pr\left(\sum_k T_k = t \mid p\right) \\ &= \binom{N}{t} p^t (1 - p)^{N-t}.\end{aligned}$$

よって  $\sum_k x_k = t$  のとき、

$$\begin{aligned}\Pr(X_1 = x_1, \dots, X_N = x_N \mid T(X_1, \dots, X_N) = t, p) &= \frac{p^t (1 - p)^{N-t}}{\binom{N}{t} p^t (1 - p)^{N-t}} \\ &= \frac{1}{\binom{N}{t}}.\end{aligned}$$

これは  $p$  に依存していない。

## 十分統計量の例 2/3

$X_1, X_2$  が独立同分布なポアソン分布

$$\Pr(X_1 = x_1, X_2 = x_2 \mid \lambda) = \frac{\lambda^{x_1+x_2}}{x_1!x_2!} e^{-2\lambda}$$

に従っているとき、 $T(x_1, x_2) = x_1 + x_2$  は  $\lambda$  の十分統計量である。実際、

$$\begin{aligned}\Pr(T(X_1, X_2) = t \mid \lambda) &= \sum_{x=0}^t \Pr(X_1 = x, X_2 = t-x \mid \lambda) \\ &= \sum_{x=0}^t \frac{\lambda^t}{x!(t-x)!} e^{-2\lambda} \\ &= \frac{1}{t!} (2\lambda)^t e^{-2\lambda}.\end{aligned}$$

よって  $x_1 + x_2 = t$  のとき、

$$\begin{aligned}\Pr(X_1 = x_1, X_2 = x_2 \mid T(X_1, X_2) = t, \lambda) &= \frac{\lambda^{x_1+x_2}}{x_1!x_2!} e^{-2\lambda} t! (2\lambda)^{-t} e^{2\lambda} \\ &= \binom{t}{x_1} 2^{-t}.\end{aligned}$$

これは  $\lambda$  に依存してない ( $X_1 \sim \text{Binom}(t, \frac{1}{2})$ )。

## 十分統計量の例 3/3

$X$  が従っている分布が対称であるとする。つまり

$$\Pr(X \in A \mid \theta) = \Pr(-X \in A \mid \theta)$$

このとき、 $T(x) = |x|$  は  $\theta$  の十分統計量である。

任意の  $\delta > 0$  について

$$\begin{aligned}\Pr(X \in [t, t + \delta] \mid T(X) = t, \theta) &= \lim_{\epsilon \rightarrow 0} \frac{\Pr(X \in [t, t + \delta], |X| \in [t, t + \epsilon] \mid \theta)}{\Pr(|X| \in [t, t + \epsilon] \mid \theta)} \\&= \lim_{\epsilon \rightarrow 0} \frac{\Pr(-X \in [t, t + \delta], |X| \in [t, t + \epsilon] \mid \theta)}{\Pr(|X| \in [t, t + \epsilon] \mid \theta)} \\&= \lim_{\epsilon \rightarrow 0} \frac{\Pr(X \in [-t - \delta, -t], |X| \in [t, t + \epsilon] \mid \theta)}{\Pr(|X| \in [t, t + \epsilon] \mid \theta)} \\&= \Pr(X \in [-t - \delta, -t] \mid T(X) = t, \theta)\end{aligned}$$

一方で  $\Pr(X \notin ([t, t + \delta] \cup [-t - \delta, -t]) \mid T(X) = t, \theta) = 0$  なので、よって、 $\Pr(X = t \mid T(X) = t, \theta) = \Pr(X = -t \mid T(X) = t, \theta) = \frac{1}{2}$  である。

# 分解定理 1/2

## Theorem (フィッシャー・ネイマンの分解定理)

確率分布族  $(p_\theta)_\theta$  について以下は同値。

1.  $T: \mathcal{X} \rightarrow \mathcal{Y}$  が  $(p_\theta)_\theta$  の十分統計量である。
2. ある  $u: \mathcal{X} \rightarrow \mathbb{R}$ ,  $(v_\theta: \mathcal{Y} \rightarrow \mathbb{R})_\theta$  が存在して

$$p_\theta(x) = u(x)v_\theta(T(x)).$$

離散の場合 (1  $\rightarrow$  2).

$$\begin{aligned}\Pr(X = x \mid \theta) &= \Pr(X = x, T(X) = T(x) \mid \theta) \\ &= \Pr(X = x \mid T(X) = T(x), \theta) \Pr(T(X) = T(x) \mid \theta) \\ &= u(x)v_\theta(T(x)).\end{aligned}$$



## 分解定理 2/2

### Theorem (フィッシャー・ネイマンの分解定理)

確率分布族  $(p_\theta)_\theta$  について以下は同値。

1.  $T: \mathcal{X} \rightarrow \mathcal{Y}$  が  $(p_\theta)_\theta$  の十分統計量である。
2. ある  $u: \mathcal{X} \rightarrow \mathbb{R}$ ,  $(v_\theta: \mathcal{Y} \rightarrow \mathbb{R})_\theta$  が存在して

$$p_\theta(x) = u(x)v_\theta(T(x)).$$

離散の場合 (2  $\rightarrow$  1).

$$\begin{aligned}\Pr(T(x) = t \mid \theta) &= \sum_{x \in T^{-1}(t)} \Pr(X = x \mid \theta) = \sum_{x \in T^{-1}(t)} u(x)v_\theta(T(x)) \\ &= \left( \sum_{x \in T^{-1}(t)} u(x) \right) v_\theta(t).\end{aligned}$$

$$\begin{aligned}\Pr(X = x \mid T(X) = T(x), \theta) &= \frac{u(x)v_\theta(T(x))}{\left( \sum_{x' \in T^{-1}(T(x))} u(x') \right) v_\theta(T(x))} \\ &= \frac{u(x)}{\sum_{x' \in T^{-1}(T(x))} u(x')}. \quad (\theta \text{ に依存してない})\end{aligned}$$

# 尤度比は十分統計量

## Lemma

パラメータ空間  $\Theta = \{\theta_0, \theta_1, \dots, \theta_k\}$  が有限集合とする。

$$T(x) = \left( \frac{\Pr(X = x \mid \theta_1)}{\Pr(X = x \mid \theta_0)}, \frac{\Pr(X = x \mid \theta_2)}{\Pr(X = x \mid \theta_0)}, \dots, \frac{\Pr(X = x \mid \theta_k)}{\Pr(X = x \mid \theta_0)} \right)$$

は十分統計量である。

## Proof.

$$\Pr(X = x \mid \theta_\ell) = \begin{cases} \Pr(X = x \mid \theta_0) & \text{if } \ell = 0 \\ \Pr(X = x \mid \theta_0) T(x)_\ell & \text{otherwise.} \end{cases}$$

よって

$$u(x) = \Pr(X = x \mid \theta_0), \quad v_{\theta_\ell}(t) = \begin{cases} 1 & \text{if } \ell = 0 \\ t_\ell & \text{otherwise.} \end{cases}$$

とおくと  $\Pr(X = x \mid \theta_\ell) = u(x)v_{\theta_\ell}(T(x))$ .



# 最小十分統計量

## Definition (最小十分統計量)

十分統計量  $T: \mathcal{X} \rightarrow \mathcal{Y}$  が最小十分統計量  $\stackrel{\text{def}}{\iff}$  任意の十分統計量  $U: \mathcal{X} \rightarrow \mathcal{Z}$  についてある関数  $V: \mathcal{Z} \rightarrow \mathcal{Y}$  が存在して  $T(x) = V(U(x))$ .

## Theorem

パラメータ空間  $\Theta = \{\theta_0, \theta_1, \dots, \theta_k\}$  が有限集合とする。

$$T(x) = \left( \frac{\Pr(X = x \mid \theta_1)}{\Pr(X = x \mid \theta_0)}, \frac{\Pr(X = x \mid \theta_2)}{\Pr(X = x \mid \theta_0)}, \dots, \frac{\Pr(X = x \mid \theta_k)}{\Pr(X = x \mid \theta_0)} \right)$$

は最小十分統計量である。

## Proof.

$U: \mathcal{X} \rightarrow \mathcal{Z}$  を十分統計量とする。分解定理より、ある  $u$  と  $v_\theta$  が存在して

$$\Pr(X = x \mid \theta_\ell) = u(x)v_{\theta_\ell}(U(x)) \quad \forall \ell \in \{0, 1, \dots, \ell\}.$$

このとき、

$$\frac{\Pr(X = x \mid \theta_\ell)}{\Pr(X = x \mid \theta_0)} = \frac{v_{\theta_\ell}(U(x))}{v_{\theta_0}(U(x))} \quad \forall \ell \in \{1, 2, \dots, \ell\}.$$

よって  $T(x) = V(U(x))$  と表すことができる。

# ラオブラックウェルの定理

## Theorem (ラオブラックウェルの定理)

損失関数  $L(\theta, d)$  が  $d$  について **凸関数** と仮定する。任意の推定関数  $\delta: \mathcal{X} \rightarrow \mathbb{R}$  と十分統計量  $T: \mathcal{X} \rightarrow \mathcal{Y}$  についてある (ランダムでない) 推定関数  $\eta: \mathcal{Y} \rightarrow \mathbb{R}$  が存在して、 $R(\theta, \eta \circ T) \leq R(\theta, \delta)$  である。

## Proof.

$$\eta(t) = \mathbb{E}[\delta(X) \mid T(X) = t]$$

とおく。このとき、

$$\begin{aligned} R(\theta, \eta \circ T) &= \sum_t \Pr(T(X) = t \mid \theta) L(\theta, \eta(t)) \\ &= \sum_t \Pr(T(X) = t \mid \theta) L(\theta, \mathbb{E}[\delta(X) \mid T(X) = t]) \\ &\leq \sum_t \Pr(T(X) = t \mid \theta) \mathbb{E}[L(\theta, \delta(X)) \mid T(X) = t] \quad (\text{イエンセン}) \\ &= \mathbb{E}[L(\theta, \delta(X)) \mid \theta] = R(\theta, \delta). \end{aligned}$$





# 決定的な推定

## Lemma (脱ランダム化)

損失関数  $L(\theta, d)$  が  $d$  について **凸関数** と仮定する。任意の (ランダムでもよい) 推定関数  $\delta: \mathcal{X} \rightarrow \mathbb{R}$  についてあるランダムでない推定関数  $\delta': \mathcal{X} \rightarrow \mathbb{R}$  が存在して、 $R(\theta, \delta') \leq R(\theta, \delta)$  である。

## Proof.

$\delta'$  を以下で定義する。

$$\delta'(x) = \mathbb{E}[\delta(x)]$$

ここで  $\mathbb{E}[\cdot]$  は推定関数の持つランダムネスについての期待値である。目的の不等式はイェンセンの不等式により得られる。  $\square$

## 課題

以下の問に答えよ。

1.

$$p(\mathbf{x} \mid \theta) = \frac{1}{\sqrt{(2\pi)^N}} e^{-\sum_{k=1}^N \frac{(x_k - \theta)^2}{2}}$$

のとき、以下の  $T: \mathbb{R}^N \rightarrow \mathbb{R}$  が十分統計量かどうか示せ。結論だけでなく理由も示すこと。

$$T(\mathbf{x}) = \sum_{k=1}^N x_k.$$

2.

$$\Pr(X_1 = x_1, \dots, X_N = x_N \mid \lambda) = \frac{1}{\prod_{k=1}^N x_k!} \lambda^{\sum_{k=1}^N x_k} e^{-N\lambda}$$

のとき、 $\lambda$  の推定関数  $\delta(\mathbf{x}) = x_1$  と十分統計量  $T(\mathbf{x}) = \sum_{k=1}^N x_k$  を用いてラオ-ブラックウェルの定理に基づいて、よりよい推定関数  $\eta \circ T$  をもとめよ。